

1634

image -

**PATENT APPLICATION**

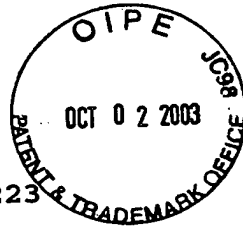
**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:

**Preben LEXOW**

Appln. No.: **09/886,223**

Filed: **June 22, 2001**



Conf. No.: **5556**

Group Art Unit: **1634**

Examiner: **Whisenant, E.**

For: **SEQUENCING METHOD USING MAGNIFYING TAGS**

**SUBMISSION OF PRIORITY DOCUMENTS**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

Submitted herewith is the certified copy for each of the priority documents (i.e., NO 19986133, NO 19996330, NO 19996331, NO 19996332, NO 19996333, NO 19996334, NO 19996335, NO 19996336, NO 19996337, NO 19996338 and NO 19996339) on which claim to priority was made under 35 U.S.C. § 119. The Examiner is respectfully requested to acknowledge receipt of said priority documents.

Respectfully submitted,

**SUGHRUE MION, PLLC**

Telephone: (202) 293-7060

Facsimile: (202) 293-7860

WASHINGTON OFFICE

**23373**

CUSTOMER NUMBER

Gordon Kit  
Registration No. 30,764

Date: **October 2, 2003**



KONGERIKET NORGE  
The Kingdom of Norway

Bekreftelse på patentsøknad nr  
*Certification of patent application no*

1998 6133

► Det bekreftes herved at vedheftede dokument er nøyaktig utskrift/kopi av ovennevnte søknad, som opprinnelig inngitt 1998.12.23

► *It is hereby certified that the annexed document is a true copy of the above-mentioned application, as originally filed on 1998.12.23*

2003.07.14

*Freddy Strømmen*

Freddy Strømmen  
Seksjonsleder

*Line Reum*

Line Reum



**PATENTSTYRET**<sup>®</sup>  
Styret for det industrielle rettsvern

16

PATENTSTYRET

23.DES98 986133

**Fra:** Preben Lexow

**Til:** Patentstyret  
- styret for det industrielle rettsvern

**Emne:** Søknad om patent

**Dato:** 23/12-1998

**Ant.sider (ekskl. denne):** 12

## Innledning

Min oppfinnelse angår en helt ny innfallsvinkel til en ikke-gelbasert DNA-sekvenseringsmetode. Helt siden Watson og Crick klarte DNA-molekylets struktur i 1953 har genforskere ønsket å finne raske og billige måter å sekvensere DNA-molekyler. Den første metoden brukte samme prinsippet som for protein-sekvensering: molekylet ble fragmentert i mindre biter, base sammensetningen for hver bit kartlagt og ved å finne overlappende sekvenser ble den opprinnelige sekvensen satt sammen. Metoden hadde store begrensninger både med tanke på størrelsen av DNA-molekylene som kunne sekvenseres og tiden som ble brukt. Det var derfor et stort gjennombrudd da Sanger/Barrell og Maxam/Gilbert utviklet to nye metoder for DNA-sekvensering i årene 1975-77. Alle metoder som brukes i større utstrekning i dag er basert på Sanger/Barrells metode og utviklingen innenfor DNA-sekvensering de siste 23 årene har stort sett vært modifiseringer av denne.

I 1988 kom imidlertid DNA-sekvenseringsteknologi i et helt nytt fokus. Ledet an av USA gikk 18 land sammen om det kanskje største enkeltprosjektet i vitenskapshistorien, sekvenseringen av hele det humane genom på  $3 \times 10^9$  bp («the Human Genom Project»), pluss flere andre mindre genomer. Per i dag er målsettingen å bli ferdig i løpet av år 2005. Til tross for at prosjektet binder opp store vitenskapelige ressurser og får en prislapp på omlag 9 milliarder \$ regnes utbyttet av prosjektet som viktig nok til at prisen kan forsvares.

En viktig del av prosjektet er å utvikle nye metoder for DNA-sekvensering som er både rimeligere og raskere enn dagens teknologi. I prinsippet kan disse inndeles i gelbaserte- (stort sett nye varianter av Sanger/Barrell metoden) og ikke-gelbaserte teknikker. De ikke-gelbaserte teknikkene har sannsynligvis det største potensialet og massespektrometri, flowcytometri og bruk av «genchips» som hybridiserer små DNA-molekyler er noen av innfallsvinklene som er under utprøving. Skulle man lykkes i å finne metoder som er vesentlig bedre enn dagens vil det resultere i en revolusjon ikke bare for genforskningen men også for moderne medisin da man får muligheten til omfattende gentesting på pasienter. Det økonomiske potensialet i en slik metode er naturligvis meget stort.

## Metoden

### **Grunnprinsippet**

Som illustrert i fig. 1 kan metoden inndeles i fire trinn:

Første trinn tar utgangspunkt i en ren DNA-populasjon bestående av DNA-sekvensen som skal sekvenseres. DNA-molekylene kuttet/brekkes på en uspesifikk måte slik at det dannes en populasjon med DNA-molekyler bestående av biter (heretter kalt DNA-biter) av den opprinnelige sekvensen.

Annet trinn består i å erstatte baseparene i DNA-bitene med 4 ulike DNA-sekvenser (heretter kalt DNA-fragmenter) som representerer hver av de fire basene Adenin, Cytosin, Guanin og Thymin. Der hvor det har vært baspar A-T settes det altså inn «fragment A», C-G byttes ut med «fragment C», osv. Dermed genereres nye DNA-molekyler hvor den opprinnelige baserekkefølgen på f.eks. ACGTT... erstattes med fragment A - fragment C - fragment G osv. Lengden på disse fire DNA-fragmentene kan i prinsippet variere i lengde fra 2bp til flere hundre kbp (eller mer om ønskelig), alt etter behov. Tilsvarende kan DNA-fragmentene inneholde reportergener og annen biologisk informasjon, eller kun bestå av sekvenser uten kjent biologisk funksjon.

I trede trinn avleses rekkefølgen på de fire typene DNA-fragmenter for hvert enkelt DNA-molekyl. Dermed finner man baserekkefølgen i de opprinnelige DNA-bitene indirekte.

I fjerde trinn benytter et dataprogram overlappene mellom DNA-bitene til å sette sammen informasjonen fra trinn 3 til sekvensen på DNA-sekvensene som ble brukt som utgangspunkt.

### **Trinn 1**

Første trinn i metoden består i å behandle en liten mengde DNA med DNase, sonikering eller lignende teknikker slik at DNA-molekylene blir fragmentert i biter (kalt DNA-biter). Ved å justere parametrene i disse teknikkene har man muligheten til å justere den gjennomsnittlige størrelsen på DNA-bitene (det optimale vil som regel være å ha gjennomsnittlige størrelser på noen hundre basepar). Metodene bør i tillegg være relativt uspesifikke med tanke på hvor de kutter/brekker DNA-molekylene slik at man statistisk sett vil få DNA-biter som er kuttet/brukket på de fleste steder i den opprinnelige sekvensen.

### **Trinn 2**

I prinsippet kan dette trinnet løses på svært mange måter, nedenfor følger 1 alternativ:

Figur 2 illustrerer en metode som baserer seg på restriksjonsenzymer som kutter utenfor sitt eget DNA-bindingssete. Disse enzymerne viser ingen spesifisitet for sekvensen som kuttet og de kan derfor generere overheng med alle typer basesamensetning. For å illustrere metoden er det brukt et enzym som kun genererer overheng på 1bp. I praksis bør man nok velge enzymer som genererer 3-4bp overheng. Ved å bruke termostabile ligaser vil spesifisiteten likevel være ivarettatt mens effektiviteten øker dramatisk slik at inkubasjonstiden reduseres. Samtidig reduserer man antallet sykluser 3-4 ganger.

Populasjon med  
DNA-molekyler  
som ønskes  
sekvensert

ACGTTGACTGAA      ACGTTGACTGAA  
ACGTTGACTGAA      ACGTTGACTGAA  
ACGTTGACTGAA      ACGTTGACTGAA

### Trinn 1 (fragmentering)



ACGTT   GACTGAA      ACGT   TGACTGAA  
ACGTTGA   CTGAA      ACGTT   GACTGAA  
AC   OTTGACTGAA      ACGTTGACTG   AA

### Trinn 2 (konvertering)



Fragment A | Fragment C | Fragment G | Fragment T | Fragment T      Fragment G | Fragment A | Fragment C | Fragment T  
Fragment A | Fragment C | Fragment G | Fragment T | Fragment T | Fragment G | Fragment A      Fragment C | Fragment T  
Fragment A | Fragment C      Fragment G | Fragment T | Fragment T | Fragment G | Fragment A | Fragment C | Fragment T



### Trinn 3 (avlesning):

Rekkefølgen på fragmentene «A», «C», «G» og «T»  
avleses for hvert DNA-molekyl.



### Trinn 4 (rekonstruksjon):

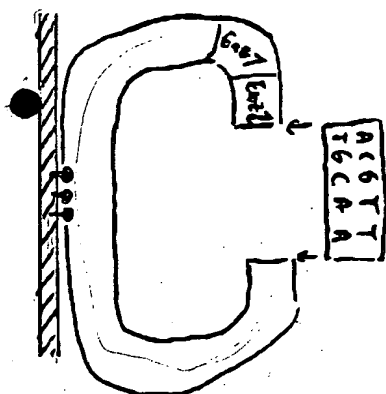
Et dataprogram setter DNA-bitene  
sammen til den opprinnelige sekvensen:

AC  
ACGT  
ACGTT  
ACGTT  
ACGTTGA  
ACGTTGACTG  
GTTGACTGAA  
TGACTGAA  
GACTGAA  
GACTGAA  
CTGAA  
AA

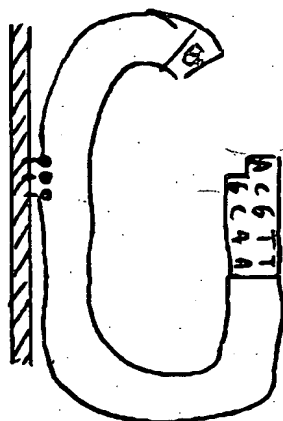
ACGTTGACTGAA

Fig.1. Grunnprinsippet for DNA-sekvenseringsmetoden. Størrelsen på DNA-bitene vil naturlig nok være større enn angitt i denne figuren

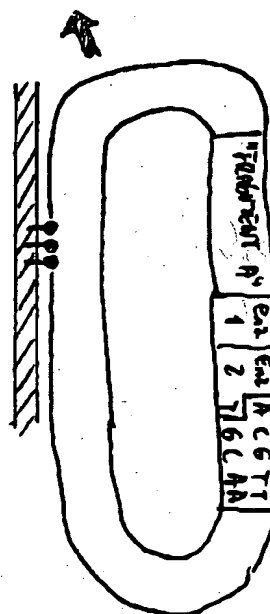
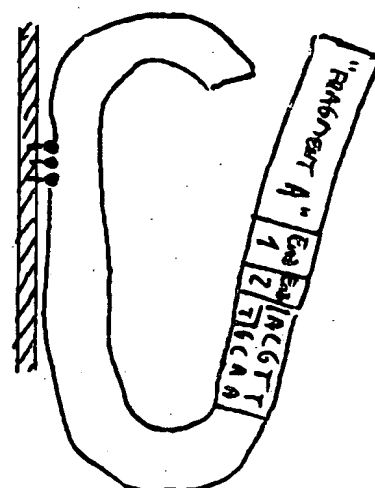
# 1) INNLIGERING AV "DNA-BITER"



# 2) GENERERING AV OVERHENG V.H.A. ENZ 1 + ENZ 2



# 3) INNLIGERING AV "DNA-FRAGMENTER"



# 4) BLUNT END-LIGERING

Fig.2 1) DNA-bitene fra Trinn 1 liggeres inn i et plasmid som har bindingssteder for et restriksjonsenzym ("Enz 1") som genererer blunt-end kutt og et som kutter utenfor sitt eget DNA-bindingssete og genererer 1bp overheng ("Enz 2"). I tillegg er det innkorporert biotinbaser i plasmidet slik at det henger fast i det streptavidinbehandlede reagensrøret hvor reaksjonen foregår. Fordelen med sistnevnte er at biprodukter fra enzymreaksjonene kan fjernes for hvert trinn. 2) reagensrøret skylles og en ny reaksjonsblanding inneholdende "Enz 1" og "Enz 2" tilsettes og inkuberes slik at det dannes en blunt ende og en ende hvor den første basen i DNA-biten utgjør et overheng. 3) reagensrøret skylles på nytt og en reaksjonsblanding inneholdende fire ulike DNA-fragmenter sammen med en termostabil DNA-ligase (f.eks. Pfu eller Taq DNA-ligase) tilsettes og inkuberes. Fordelen med de termostabile ligasene er at de liggerer meget spesifikt samtidig som de ikke liggerer blunt ender. Dermed vil «fragment A» liggeres inn der hvor det er adenosin som overheng; «fragment C» der hvor det er en cytosin, osv. 4) reagensrøret skylles nok en gang og en reaksjonsblanding med T4 ligase tilsettes og inkuberes slik at blunt endene liggeres. Siden de innligerte fragmentene har seter for Enz 1 og Enz 2 slik som plasmidet hadde opprinnelig er man tilbake til utgangspunktet slik at neste base kan byttes ut med et DNA-fragment i en ny syklus.

### Trinn 3

På samme måte som for trinn 2 kan også trinn 3 utføres på mange alternative måter:

1) For å finne en bedre metode til å lokalisere fluorescerende DNA-prober brukte Weier et al (ref) en teknikk kalt «molecular combing»; En løsning med mål-DNA som probene ble hybridisert til ble plassert på en flat glassoverflate preparert slik at DNA-molekylene festet seg med en av endene til glassplaten. Deretter fikk de DNA-molekylene til å rette seg ut ved hjelp av en væskestrøm. Ved hjelp av et fluorescensmikroskop kunne de dermed observere probenes relative posisjoner på de utstrakte DNA-molekylene. Ved å bruke fire prober merket med ulike fluoroforer og som hybridiserer til de fire DNA-fragmentene som representerer «A», «C», «G» og «T» vil den ovenstående teknikken kunne brukes til å avlese sekvensrekkefølgen direkte med et fluorescensmikroskop. Ved å utvikle programvare som gjør at mikroskopet «scanner» glassplaten samtidig som sekvensrekkefølgene analyseres automatisk vil det sannsynligvis bli mulig å avlese flere tusen basepar per minutt.

2) Potensialet for hurtig avlesning vil sannsynligvis være enda større hvis man anvender et flowcytometer for å avlese de fluorescerende probene. En forutsetning for dette er at DNA-molekylene passerer avlesningsenheten på et flowcytometer i utstrukt form slik at DNA-fragmentene som representerer «A», «C», «G» og «T» vil passere i rekkefølge.

En metode for å få DNA-molekylene utstrukt er å blande dem med små glasskuler (binder DNA-molekyler naturlig) i stort overskudd slik at de binder DNA-molekylene i forholdet 1:1. DNA-molekylene vil ha mindre motstand enn glasskulene i en væskestrøm slik at de tenderer til å bevege seg bort fra hverandre helt til DNA-molekylet er utstrukt. Hvis væskestrømmen er kraftig eller glasskulene store slik at forskjellen i motstand for DNA-molekylene og glasskulene er stor vil DNA-molekylet kunne ryke. Dette problemet vil imidlertid kunne unngås ved å senke flow-hastigheten eller bruke mindre glasskuler.

En alternativ strategi er å bruke et elektrisk eller magnetisk felt i stedet for væskestrøm til å trekke partiklene forbi fluorescensdetektoren. Det kan gjøres ved å utnytte at glasskulene har positiv ladning mens DNA-molekyler er negativt ladd, eller bruke superparamagnetiske kuler istedenfor glass. Dermed vil kulene trekke DNA-molekylene som henger etter som rette tråder.

En kritisk parameter ved denne strategien er flowcytometrenes nedre deteksjonsgrenser for fluorescens. Flere grupper har klart å detektere enkelte fluorofor-molekyler ved å senke flow-hastigheten. Ønsker man imidlertid å bruke konvensjonelle flow-cytometre med analysehastigheter på 20-30.000 partikler per sekund må man bruke lengre prober slik at det kan festes mange fluoroforer til hver probe.

De raskeste flowcytometrene per i dag har kapasitet til å analysere ca. 200.000 fluorescerende partikler per sekund. Disse flowcytometrene er imidlertid ikke kommersielt tilgjengelige. I tillegg kommer at det er usikkert hvor høye flow-hastigheter DNA-molekyler tåler i utstrukt form før de ryker. Det er imidlertid realistisk å anta at DNA-molekylene vil tåle hastigheter som gjør det mulig å avlese et sted mellom 5-50 tusen basepar per sekund (med en hastighet på 50.000 bp/s vil ett flowcytometer kunne avlese 3 billioner bp (tilsvarende det humane genom) på i underkant av 17 timer. Sammenlignet med HUGO-prosjektet hvor man regner med å bruke 17 år og betydelige laboratorieressurser i over 20 land på å få sekvensert det første humane genom vil dette representere en betydelig nyvinning.)

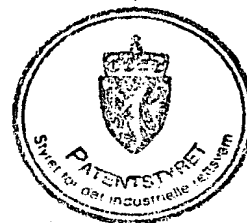
3) En rekke andre strategier kan tenkes for å avlese rekkefølgen på DNA-fragmentene i de ulike DNA-bitene. I tillegg kommer at ideen om å konvertere et DNA-molekyl til DNA-fragmenter innsatt i samme rekkefølge som de opprinnelige baseparene vil kunne løse en del av de problemene som har oppstått i forbindelse med andre DNA-sekvenseringsprosjekter.

#### Trinn 4

I siste trinn brukes et dataprogram til å sette sammen sekvensbitene til den endelige sekvensen. Sannsynligheten for at det skal inntreffe feil i dette trinnet avhenger hovedsakelig av 5 parametre: Lengden på DNA-molekylet som skal sekvenseres, hvor randomisert baseparsammensetningen av DNA-sekvensen er, lengden på DNA-bitene som skal avleses, antallet DNA-biter som blir avlest, feilfrekvensen i trinn 2 og 3.

Jeg har allerede laget et dataprogram for å analysere betydningen av de ovenstående parametrene. Med utgangspunkt i human genomisk DNA som allerede er sekvensert viser analysene at med en DNA-bit lengde på 30 fragmenter, avlesning av  $6 \cdot 10^6$  DNA-biter og en feilfrekvens i trinn 2 og 3 på 10% (med tanke på punktmutasjoner) vil et humant genom kunne avleses i en sekvensreaksjon og med svært få punktmutasjoner/delesjoner. Et unntak er imidlertid svært ikke-randomiserte områder (satellitt DNA og andre repetitive områder) hvor DNA-bit lengdene må økes. Den biologiske informasjonen i disse områdene er imidlertid av underordnet betydning sammenlignet med kodende sekvenser og cis-regulatoriske elementer.

Dataanalysene viser også at selv en meget høy feilfrekvens i trinn 2 og 3 kompenseres ved at DNA-bitene avleses mange ganger. Ved å avlese f.eks. 10 ganger så mange basepar som sekvensens lengde vil man eliminere de fleste delesjoner og punktmutasjoner selv med en meget høy feilfrekvens i trinn 2 og 3.



### Patentkrav

Fremgangsmåte for DNA-sekvensering som består av fire trinn:

Første trinn tar utgangspunkt i en ren DNA-populasjon bestående av DNA-sekvensen som skal sekvenseres. DNA-molekylene kuttet/brekkes på en uspesifikk måte slik at det dannes en populasjon med DNA-molekyler bestående av biter (heretter kalt DNA-biter) av den opprinnelige sekvensen.

Annet trinn består i å erstatte baseparene i DNA-bitene med 4 ulike DNA-sekvenser (heretter kalt DNA-fragmenter) som representerer hver av de fire basene Adenin, Cytosin, Guanin og Thymin. Der hvor det har vært baspar A-T settes det altså inn «fragment A», C-G byttes ut med «fragment C», osv. Dermed genereres nye DNA-molekyler hvor den opprinnelige baserekkefølgen på f.eks. ACGTT... erstattes med fragment A - fragment C - fragment G osv. Lengden på disse fire DNA-fragmentene kan i prinsippet variere i lengde fra 2bp til flere hundre kbp (eller mer om ønskelig), alt etter behov. Tilsvarende kan DNA-fragmentene inneholde reportergener og annen biologisk informasjon, eller kun bestå av sekvenser uten kjent biologisk funksjon.

I tredje trinn avleses rekkefølgen på de fire typene DNA-fragmenter for hvert enkelt DNA-molekyl. Dermed finner man baserekkefølgen i de opprinnelige DNA-bitene indirekte.

I fjerde trinn benytter et dataprogram overlappene mellom DNA-bitene til å sette sammen informasjonen fra trinn 3 til sekvensen på DNA-sekvensene som ble brukt som utgangspunkt.



**Sammendrag**

Min oppfinnelse angår en helt ny innfalsvinkel til en ikke-gelbasert fremgangsmåte for DNA-sekvensering. Fremgangsmåten består av flere trinn og det finnes flere alternative løsninger for hvert av dem. Det sentrale er imidlertid at DNA-sekvensen «forstørres» ved å erstatte basepar med lengre DNA-sekvenser. Dermed åpnes mange muligheter for avlesning av DNA-sekvensen.

